PARTIAL TRANSLATION OF JP 58(1983)-193595 A

Publication Date: November 11, 1983
Title of the Invention: TELEPHONE INFORMATION INPUT APPARATUS
Patent Application Number: 57-75282
Filing Date: May 7, 1982
Inventor: K. NAKAMURA
Applicant: HITACHI, LTD.

Claim

1. A telephone information input apparatus characterized in that, in a speech recognition apparatus having a plurality of sets of phoneme standard patterns classified for each speaker and a phoneme sequence word dictionary corresponding to words to be recognized, a first pseudo phoneme pattern for detecting the presence/absence of a push button signal is contained in a particular set of phoneme standard patterns, a second pseudo phoneme pattern for recognizing each push button signal is contained in the remaining sets, and a pseudo phoneme sequence word dictionary corresponding to the first and second pseudo phoneme patterns is provided.

(Page 1, right column, lines 3-10)

According to the push button (PB) signal input 1), among artificial signals obtained by combining speech band sine wave 2 frequencies (one high frequency and one low frequency), a combination of four low frequencies and four high frequencies is currently used with the specification unified, and in principle, 16 kinds of information can be inputted (see Matsuzaka, Uehara, Yatani: A signal system for a push button dial telephone: Nippon Telegraph and Telephone Public Corporation, Telecommunication Research Institute, Research Commercialization Report 17-11, p. 241, November, 1968).

THIS PAGE BLANK (USPTO)

(Page 3, upper right column, line 8 – lower left column, line 12)

The present invention is characterized in that a PB signal is given as one phoneme and word to the speech recognition part 13 shown in FIG. 1 as described above, and the PB signal is detected in the same manner as that of speech recognition, whereby the speech and the PB signal may be used together as information input means. The coexistence of the speech signal and the PB signal is not assumed.

First, simply, the case where frame-based phoneme recognition is performed on a round-robin system with respect to all 16 sets of phoneme patterns will be considered. One PB signal for each of 16 sets (i.e., 16 kinds of PB signals in total) is assigned as a pseudo phoneme standard, and a feature pattern required for the detection thereof may be stored in a phoneme standard pattern memory. The word dictionary corresponding to the PB signals may be configured so as to satisfy the condition that a pseudo phoneme standard with respect to the same PB signal is maintained for a period of time or longer (e.g., 40 milliseconds or longer according the current stipulation) which is required for the reception and detection.

Next, in the case of hierarchical processing in which first-stage recognition is performed by the first two representative clusters, the following is considered.

1) It is detected that a signal is a PB signal in the first-stage recognition.

2) In the case where it is detected that a signal is a PB signal in the first-stage recognition, it is recognized which PB signal is detected in second-stage recognition.

THIS PAGE BLANK (USPTO)

⑲ 日本国特許庁（JP）　　⑪ 特許出願公開

⑫ 公開特許公報（A）　　昭58—193595

⑤1Int. Cl.³	識別記号	庁内整理番号
G 10 L 1/00		7350—5D
H 04 M 1/26		7251—5K
11/00		7345—5K

⑬公開　昭和58年(1983)11月11日

発明の数　1
審査請求　未請求

（全 5 頁）

⑤4電話情報入力装置

⑦1出　願　人　株式会社日立製作所
東京都千代田区丸の内1丁目5番1号
⑦4代　理　人　弁理士　薄田利幸

⑦1特　願　昭57—75282
⑦2出　願　昭57(1982)5月7日
⑦2発　明　者　中田和男
国分寺市東恋ケ窪1丁目280番地株式会社日立製作所中央研究所内

明　細　書

発明の名称　電話情報入力装置

特許請求の範囲

1．話者別に分類された複数組の音素標準パタンと認識すべき単語に対応した音素系列単語辞書とを有する音声認識装置において、押しボタン信号の有無を検出するための第1の擬似音素パタンを音素標準パタンの特定の組の中に有し、残りの組内に個々の押ボタン信号を認識するための第2の擬似音素パタンを持ち、第1および第2の擬似音素パタンに対応した擬似音素系列単語辞書を設けたことを特徴とする電話情報入力装置。

発明の詳細な説明

本発明は電話による情報の入力、とくに音声認識を利用した情報入力装置に係り、特にその機能を押しボタン信号による入力の併用にも拡大するのに好適な音声認識装置の構成に関する。

従来の電話機による情報の入力手段には次の2つがある。1）押しボタン信号入力（以下ＰＢ入力と略す）、2）音声認識入力（以下音声入力と略す）。

1）は音声帯域正弦波2周波（高域，低域各1周波）の組み合せによる人工的な信号で、現在規格を統一されて使用されているものは低域4周波、高域4周波の組み合せで原理的に１６種類の情報を入力することができる（松坂，上原，矢谷：押しボタンダイヤル電話用信号方式：日本電信電話公社電気通信研究所研究実用化報告１７－１１，Ｐ２４１昭和４３年１１月参照）。

この1）の方法によれば情報は確実に入力できるが、情報をすべて数字コードに変換して入力しなければならず、また押しボタン電話機が使えないところでは情報を入力することができない。

2）は音声認識によつて、音声のままで情報を直接入力しようとするもので、便利ではあるが、常に確実，正確に情報が入力できるとは限らない（長島，中津：音韻単位の標準パタンを用いた実時間単語音声認識装置、日本音響学会音声研究会資料、８７８－２２，１９７９，渡辺，互理，千葉

BNSDOCID: <JP_____358193595A_I_>

他；不特定話者用音声認識装置ＳＲ－１０００シリーズ，日本音響学会講演論文集，３－１－２４，１９８１年５月）

本発明の目的は、従来の音声認識装置の構成を基本とし、これにごくわずかの追加を行うことによつて、あらかじめ音声信号かＰＢ信号かがわからなくても、それぞれ認識が行なわれ、そのことによつて音声とＰＢ信号を自由に併用して使用でき、電話機による情報入力の機能を拡大する手段を提供することにある。

まず、従来の電話情報入力用音声認識システムの構成を第１図に示す。

第１図において、加入者電話機１１から交換機１２を通つた音声信号１２１は音声認識部１３に入力され、業務処理部１４からの認識要求信号１４１を受けてその認識処理をおこなう。主業務処理部１４では、認識結果を確認するために認識完了信号１４２を受けて音声出力部１５に出力要求信号１５１を送出し、音声出力の終了を出力完了信号１５２により確認する。

一方、交換機１２からの応答信号１２２を受けて発信制御部１６から出力された応答信号１６１が主業務処理部１４に入力されると、電文処理部１７にたいして送信要求信号１４３が送出される。

これを受けた電文処理部１７は通信制御部１８にたいし、送信要求信号１７１を送ることによりリレーコンピュータ１９から発せられ通信制御部１８を通つた電文１８１を受信して発信制御部１６にたいし発信要求信号１７２を送り信号162を発信させる。

第２図は第１図における音声認識部１３のブロック構成を示す。

第２図(b)で示す波形の入力音声２０（ichi）から音声分析部２１において抽出された特徴パラメータの系列２１１と音素標準パターンメモリ２２中に格納されている例えば１６組の音素（a，i，……など）の特徴パラメータ（最尤スペクトルパラメータ、ＬＰＣケプストラム係数など）との距離が距離計算部２３において計算される。

距離計算部２３から出力された距離の系列231と単語辞書メモリ２４中に格納されている標準単語（たとえば、単語番号１，２，……に対応してそれぞれ音素記号系列ｉ・ｃｈｉ，ｎｉ，……などで表わされる単語）との非線形マッチング演算がＤＰマッチング部２５においておこなわれ、その結果得られた距離和２５１の大きさにもとづいて単語判定部２６で入力音声の判定がおこなわれ、認識結果２７が出力される。

この認識処理の特徴は、電話入力された不特定話者の音声認識を、１６組の音素標準パターンによるフレーム別認識をおこなう第１段と、フレーム別認識の結果と音素記号系列単語辞書とのＤＰマッチングをおこなう第２段とからなる２段のパタン照合に分解し、第１段では音素標準パタンにたいして話者の音声波形における音響的な特性にもとづいて１６組のクラスタリング（組み合け）をおこない、第２段では１つの単語に対して複数個の音素記号系列単語辞書をもうけて、発話の変化、たとえば母音の無声化や鼻音化、に対処していることである。本方式はこの２段処理によつて所要メモリ量が少なくかつ不特定話者の音声に対して暖い認識能力を持つことが知られている。

「一桁の数字音（０～９の１０個）」および、「はい」，「いいえ」，「どうぞ」，「もう一度」，「保りゆう（保留）」，「とりけし（取り消し）」の６語を含む１６語に対して６２０名の男女による認識結果の一例を表１に示す（電電公社通信研究所発表）。

表１　男女別の誤り率〔%〕

	男	女	平均
尤度による距離	4.45	3.22	3.86
ＬＰＣケプストラム距離	4.57	3.18	3.90

なお、表１で距離尺度としてとられているのは、音声認識のための特徴として使われるパラメータの一例であり、このいづれを用いても誤り率はほとんど差のないことをあらわしている。

この方式のもう一つの特徴は、１６組に分類された多数の（最大４０個程度）音素標準パターンとの整合によつて、フレーム別に音素系列を認識

BNSDOCID: <JP_____358193595A__I_>

し、その結果と単語音素系列との比較によつて単語を認識するに当つて、その処理量を軽減し、実時間認識を可能にするため、そのフレーム別音素認識を第3図(a)に示すように2段に分けて階層的に行つていることである。すなわち、16組のパタンの中、男声の代表として作られている例えば第1の組と、女声の代表として作られている例えば第16の組との2組の標準パターンで、まず第1段の認識を行い、その中で整合度の良いものN個をえらび、そのN語に対象を限定し、改めて上記16組の音素標準パターンのすべてを使つて再認識を行う。Nの数としては第3図(b)に示す実験結果からN＝4にとれば、誤認識による誤りが少なく、処理量（計算量）も少なくてすむことがわかる。ここで計算量の比率とは、

$$\frac{\text{一次選択ありの計算量}}{\text{一次選択なしの計算量}} \times 100$$ で与えられる。

結果的に16語に対して16語×16語＝256組×語の処理を、2組×16語＋16組×4語＝96組×語の処理に軽減している。

さて通常のPB信号は、いわゆるPB信号受信器で検出される。通常の使い方では情報を入力する信号の形式がPB信号であるか音声信号であるかはあらかじめ決まつており、分離して行なわれる。例えば通常の使い方ではPB信号は情報センターへのアプローチに使われ、第1図における発信制御部13で受信検出される。

本発明のポイントはすでに述べた第1図の音声認識部13へ、音素および単語の1個としてPB信号を加え、音声認識と全く同じ形式でPB信号を検出することによつて、音声とPB信号を情報入力手段として併用してもよいようにしようとするものである。ただし音声信号とPB信号の同時共存は仮定しない。

まず簡単に、16組の音素パタンのすべてと総当りでフレーム別音素認識が行なわれる場合を考える。このときは16組の各組に1個、あわせて16種のPB信号を擬音素標準として割り当て、その検出に必要な特徴パターンを音素標準パターンメモリに記憶させておけばよい。PB信号に対応する単語辞書としては、受信検出しなければならないとされている時間以上（たとえば現行規定によれば40ミリ秒以上）同一のPB信号に対する擬音素標準が維持するという条件を満足するように構成すればよい。

次に最初2個の代表クラスターによつて第1段目の認識が行なわれるという階層処理の場合には次のように考える。

1) 第1段目の認識でPB信号であることを検出する。

2) 第1段目認識でPB信号と検出された場合第2段目でそのいづれであるかを認識する。

以下さらに具体的に説明する。

音韻認識において、LPC（線形予測）分析にもとづいて尤度比による整合をとる場合について考える。

0.3kHzから3.4kHzまでに帯域制限された音声信号に対して、通常p＝10次の分析が行なわれる。

この分析の結果、原理的にはp／2個のスペクトルの共振周波数いわゆるホルマント周波数が指定される。すなわちp＝10の場合、5個の周波数を指定することができる。この5個の周波数を、低，高の両周波数帯に、第4図に示すように割り当てれば、16個の周波数の中の任意の6個をカバーするように設定することができ、2組によつて任意の12個をカバーするようにすることができる。

第4図において、1，2，3，4，5は♯1クラスタの割り当て周波数を示し、Ⅰ，Ⅱ，Ⅲ，Ⅳ，Ⅴは♯1bクラスタの割り当て周波数を示す。

PB信号としては、16個の中から実際には10数字と制御用に2個（たとえば＊印と♯印）が用いられるのが普通であり、12個を検出できればよい。日本国内では低域は4周波（697，770，852，941Hz）であるが、高域は3周波（1209，1336，1497Hz）しか使つていない。

これらを検出するパラメータは次式から導出することができる。

BNSDOCID: <JP_____358193595A_I_>

指定周波数を $\{f_i\}=(f_1, f_2, f_3, f_4, f_5)$ とするとき

$$\begin{cases}\beta_i = r_i e^{j\theta_i} & r_i = e^{-\pi b_i T} \\ \overline{\beta}_i = r_i e^{-j\theta_i} & \theta_i = 2\pi f_i T\end{cases}$$

ここでTはサンプリング周期、b_i は f_i の共振帯域幅であり、PB信号の場合、許容信号周波数変動幅は±2%と規定されているから $b_i = f_i \times 4\%$ 程度にとればよい。

これから

$$(Z-\beta_1)(Z-\overline{\beta}_1)(Z-\beta_2)(Z-\overline{\beta}_2)(Z-\beta_5)(Z-\overline{\beta}_5) \quad \cdots\cdots (1)$$

の10次方程式を作り、それを

$$Z^{10}+\alpha_1 Z^9+\alpha_2 Z^8+\cdots\cdots+\alpha_9 Z+\alpha_{10} \quad \cdots\cdots (2)$$

とおいて(1)式と(2)式のZの等べき係数を $\alpha_1, \cdots\cdots \alpha_{10}$ とおけば、$(\alpha_1, \cdots\cdots \alpha_{10})$ が求められる。

音韻標準パタンとして使われる逆スペクトル係数は、このαの系列に α_0 とした1を加えた系列の相関係数として、

$$A_0 = 1+\alpha_1^2+\alpha_2^2+\cdots\cdots+\alpha_{10}^2$$
$$A_1 = \alpha_1+\alpha_1\alpha_2+\alpha_2\alpha_3+\cdots\cdots+\alpha_9\alpha_{10}$$
$$\vdots \qquad \vdots$$
$$A_9 = \alpha_9+\alpha_9\alpha_{10}$$
$$A_{10} = \alpha_{10}$$

と求められる。

#2から#15までのクラスタに記憶される音素パラメータは、現実の個々のPB信号をLPC分析することによつて求めることができる。

なお実際には12個のPB信号をすべて対象とする必要はなく、第4図にその1例を示すように、#1クラスタによつて6種類、#16クラスタによつて6種類が指定されるから、この6種類についてのみ第2段の認識実験を行えばよい。

本発明の一実施例を第2図を用いて説明する。

入力音声20（擬似音声波形としてPB信号であることもある）は音声分析部21で相関係数 $\{r_i^{(n)}\}$ の算出とLPC（線形予測）分析がされ、残差電力 $E_0^{(n)}$ が計算される。

次に距離計算部23で各フレーム毎に音素標準パタン $\{A_i^{(n)}\}$, $i=0\sim10$, $n=1\sim S$ と入力Xの相関係数 $\{r_i^{(x)}\}$, $i=0\sim10$ と E_0 によつて次式によつて尤度比が計算される。

$$L_n^2 = \left[A_0^n r_0 + 2\sum_{i=1}^{10} A_i^n r_i^x\right] / E_0^x \quad \cdots\cdots (3)$$

この L_n^2 を尺度とする入力音素系列マトリックスと音素記号系列単語辞書との間でDPマッチングによる整合がとられ、最適整合のものが認識結果として出力される。その場合、すでに説明したように16組の音韻クラスタにおいて、代表的な2つ、たとえば#1（男声代表）と#16（女声代表）のみを用いた第1段の認識が行なわれ、候補単語がN個にしぼられる。このとき、#1と#16のクラスターに追加されたPB信号検出用のパタンによつて第1候補がPB信号であると検出されたときは、N個の候補として、12種類のPB信号の中の6個を候補として第2段目の認識を行う。その他は従来の音声認識と全く同じである。

この場合、個別PB信号に対応する擬似音素パタンを2組から15組に1個づつ加えないで、第17組としてPB信号用のクラスタを構成すれば、第一段目でPB信号として検出されたときは、このクラスタについてのみフレーム別音素認識を行えばよいように構成することもできる。

また一般に行なわれている単語レベルでの複数標準パタンによる音声認識においては、PB信号に対しては16組クラスター総当りで説明した一段目の認識のみでよいことは自明である。

以上説明したように本発明によれば、音声とPB信号を何ら区別することなく電話による情報入力手段として利用することができ、音声入力の簡便さとPB入力の確実さの特色を活かした情報入力が可能となる。

たとえば、音声によつては比較的長く、文脈効果の利用しやすい制御語のみを入力し、短かくて文脈効果の利用しえない数字データはPB入力と

BNSDOCID: <JP_____358193595A_I_>

するといつた使い方も可能となる。

あるいはＰＢ電話機を利用できる人には確実なＰＢ入力を、利用できない人には音声入力を使うシステムをサービスすることもできる。

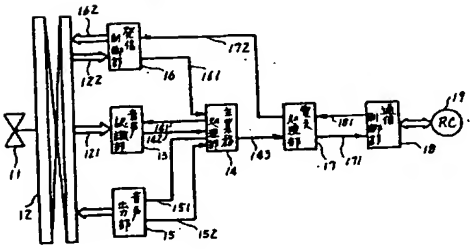
図面の簡単な説明

第１図は従来音声認識応答システムの構成図、第２図はその音声認識部の説明図、第３図は実際に行なわれている階層認識処理の説明図、第４図はＰＢ信号検出用擬似音韻パタンによる検出可能領域の説明図である。

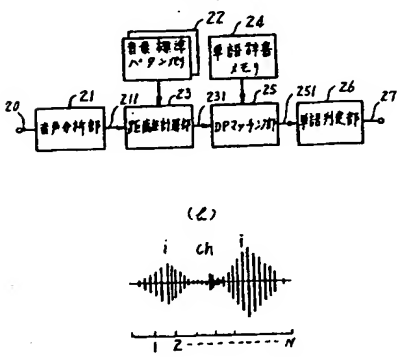
１３…音声認識部。

代理人　弁理士　薄田利幸

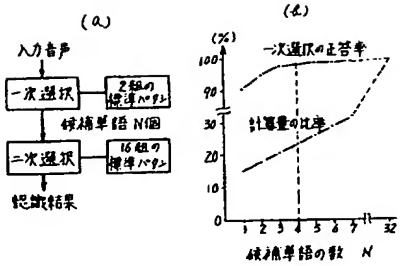
第　１　図



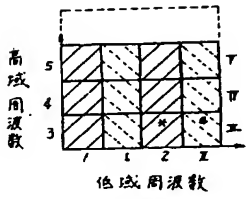
第　２　図



第　３　図



第　４　図



BNSDOCID: <JP_____358193595A__I_>

THIS PAGE BLANK (USPTO)